1. INTRODUCTION

The outbreak of the COVID-19 virus has made many changes in aspects of life in society, from the economic and educational to social aspects of experiencing the impact of this virus [1]. The government continues to seek ways to reduce the spread of the COVID-19 virus. The efforts included the Large-Scale Social Restriction (PSBB) policy [2] to the obligation of the entire community to carry out COVID-19 vaccinations to booster vaccinations. Since the first injection of the COVID-19 vaccine to President Joko Widodo at the Presidential Palace. The government continues encouraging the public to vaccinate the first and second doses of the Covid-19 vaccine. This national vaccination program was continued until the Ministry of Health (Kemenkes) through the Directorate General (Directorate General) of Disease Prevention and Control issues Circular (SE) Number HK.02.02/II/252/2022 concerning Advanced Dose (Booster) COVID-19 Vaccination [3]. Based on data from the Ministry of Health, 40.37 million booster vaccines have been given to the public as of May 3, 2022 [4].

The Circular provides many responses and opinions to the public on various social media. Although the government has socialized it, there are still many people who still do not understand the use of vaccines. In addition, many people are afraid or hesitant to be given the COVID-19 vaccine [5]. The many types of vaccines circulating in the market and the different side effects of each vaccine brand are also of the most common discussions on social media [6]. Based on data from We Are Social, the number of social media users in Indonesia will be more than 190 million. Twitter is one of the social media used by many people, which reaches 58% of the total social media users in Indonesia [7]. Due to a large number of Twitter social media users, many Indonesian people use tweets as a feature or tool to interact in disseminating information [8].

Every tweet made by the community contains sentiment. Sentiment analysis is a field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. Sentiment analysis is done to see opinions on an issue, whether they tend to be positive or negative opinions [9].

In a previous study, 2021 conducted by W. Yulita, E. D. Nugroho, and M. H. Algifari with the research topic of sentiment analysis on public opinion about the COVID-19 vaccine by implementing the Naive Bayes method. The results found that the positive sentiment was 60.3%, the neutral sentiment was 34.4%, and the negative sentiment was 5.4%. Furthermore, it produces an accuracy value of 93% from 3780 tweet data processed [10]. In a similar study in 2021 on sentiment analysis with a different method, H. Hayati and M. R. Alifi conducted a similar study. The author implements the Support Vector Machine algorithm that uses 360,000 tweet data with 120,000 data labeled positive, negative, and neutral each. This study resulted in a value of 84% of all aspects of measurement, namely accuracy, recall, f-measure, and precision [11].

In 2021, a similar study will be conducted in a study conducted by A. Baita, Y. Pristyanto, and N. Cahyono, only focusing on one type of vaccine, namely the Sinovac vaccine. The method used in this research is Support Vector Machine and K-Nearest Neighbor. With negative sentiment of 52%, much greater than the positive sentiment of 18%. Likewise, it is smaller than neutral sentiment, which has a value of 31%. This study resulted in an accuracy rate of 81% of the 221 processed data [12]. Different results are shown in the 2021 study conducted by F. Fitriana, E. Utami, and H. Al Fatta. The author applies the Support Vector Machine (SVM) and Naïve Bayes algorithms on the topic of the covid-19 vaccine. Produces the accuracy rate of 90.47% using Support Vector Machine, while the Naïve Bayes accuracy result is 88.64%. The author says that the Support Vector Machine algorithm is superior in accuracy performance, while Naïve Bayes is superior in timing performance [13].

In a study conducted by A. K. Santoso, A. Noviriandini, A. Kurniasih, B. D. Wicaksono, and A. Nuryanto, the topic of Twitter users' perceptions of COVID-19. The method used is Logistic Regression by varying the L2 and None hyperparameters. Furthermore, the results on the L2 hyperparameter obtained an accuracy value of 77% and an F1 score of 74%. Furthermore, in the hyperparameter variation None, the accuracy value is 74%, and the F1 Score is 70%. Therefore, the author says that the L2 hyperparameter value is the best variation on the Logistic Regression method [14].

In previous studies that have been carried out, the authors only used one or two classification algorithms. In this study, we will use a different method: Ensemble Stacking on the COVID-19 Vaccine Booster sentiment. The purpose of this study is to determine the level of accuracy in the Ensemble Stacking method with datasets resulting from collecting or crawling data on Twitter with the specified time and keywords and tweets taken using Indonesian-language tweets with sentiment grouping into two classes, namely positive sentiment and negative sentiment.